REMARKS

These remarks are made in response to the Office Action of October 22, 2004 (Office Action). As this response is filed after the 3-month shortened statutory period, a retroactive extension of time is herein request, and an appropriate extension fee included.

In paragraph 1 of the Office Action, the Examiner has rejected claims 1-26 under 35 U.S.C. § 102(b) as being anticipated by U.S. Patent No. 6,076,059 to Glickman, *et al.* (Glickman).

Applicants have amended independent claims 1, 6, 11, 17, and 22 to clarify that the claimed invention compares an allegedly true textual representation (a first representation) of a realization of spoken audio against a speech recognized version (second representation) of the realization. The word database is used when producing the second representation. The word database is automatically updated based on comparison results. Applicants have also amended dependent claims 2, 4, 7, 9, 12, 14, 18, 20, 22, 23, and 25 to maintain consistency with the newly amended independent claims.

These amendments are supported by page 3, lines 3-5 (automatically updating), by page 3, lines 13-14 (first representation being an allegedly true representation), by page 3, line 15 (realization is spoken audio), by page 3, lines 10-13 (second representation is a speech-recognized textual representation of the spoken audio), by page 3, lines 16-17 (speech recognition vocabulary or pronunciation data being updated), by page 4, lines 5-6 (automated vocabulary or speech recognition dictionary update process), by page 4, lines 14-16 (extends an existing vocabulary of a speech recognition system), by page 6, lines 18-21 (word database updated), between page 8, line 26 and page 9, line 6, by previous claim 5 (on a word-by-word basis), by page 1, lines 19-20, as well as by FIGS 1-5 and material contained throughout the specification.

Applicants have also added new claims 27 and 28 to emphasize various disclosed inventive aspects of the present invention that were previously unclaimed. Specifically, in claim 27, Applicants emphasize that the first representation can be obtained using an

optical character recognition technology, as supported by page 3, lines 18-19. In claim 28, Applicants have emphasized that the word database is a speaker dependent database used to adapt the speech recognition to a particular speaker, as supported by page 4, lines 5-11.

As each claim amendment is adequately supported by the specification, no new matter results from the claim amendments. Before responding to the Office Action, a brief review of the Applicants' invention may prove helpful. The Applicants' claimed and disclosed subject matter includes a system, method, and apparatus for automatically training speech recognition vocabularies that automatically extends an existing vocabulary or automatically generates a new vocabulary. The Applicants' teachings overcome conventional disadvantages of not being able to automatically update a speech recognition vocabulary based upon on large scale bodies of text, as noted at page 2, lines 14-17.

More specifically, the Applicants disclose that a speech recognition vocabulary can be used to speech-to-text convert a speech audio stream to generate a textual transcription (second representation) of the spoken text. The second representation can include non-recognized utterances, as noted at page 3, lines 14-16. The second representation can be compared against an allegedly true textual representation (first representation) of the received speech audio stream. This comparison can be used to automatically update the word database used when generating the second representation. Consequently, a speech recognition grammar can be extended or created with minimum technical effort and time and with minimal user interactions, as supported by page 3, lines 5-17.

Turning to the rejections on the art, Glickman teaches a method for aligning a text file with an audio file, where the text file includes written words and the audio file includes spoken words, there being a one-to-one correspondence between some of the written and spoken words, as noted at column 1, lines 51-55. The purpose of Glickman is

to overcome the perceived deficiencies of forced alignment (text-to-audio signal) techniques. In particular, to correct deficiencies relating to aligning text with a noisy audio stream, problems relating to gross misalignments, and/or problems relating to aligning text that does not represent the entire duration of an audio stream, as noted at column 1, lines 37-41. Glickman expressly teaches (at column 1, lines 24-26) that problems with alignment of text with audio signals are different from problems relating to speech recognition.

Referring to independent claims 1, 6, 11, 17, and 22, Applicants expressly claim that a word database is used to speech-to-text convert a realization of spoken audio into a second representation that is a speech recognized textual representation of the spoken audio. The second representation is aligned with a first representation that is an allegedly true textual representation of the spoken audio. Based upon alignment results, the word database can be automatically updated to extend the word database.

Glickman fails to expressly or inherently teach a method where a word database is extended based upon alignment results. Consequently, Applicants respectfully request the Examiner withdraw the § 102(b) rejections to the independent claims (claims 1, 6, 11, 17, and 22) as well as claims dependent upon them (claims 2-5, claims 7-10, claims 12-16, claims 18-21, and claims 23-28).

Although the current claims (as amended) that are contained herein should now be in an allowable state, Applicants shall briefly take a moment to emphasize the significance of extending the word database as claimed by the Applicants and to show how Glickman's teachings, purpose, and principle of operation are directly opposed to the Applicants' claimed system and methods.

The Applicants have claimed and disclosed a speech recognition method that overcomes the previous prior art deficiencies relating to extending speech recognition system vocabularies automatically. The updates result from results achieved when aligning an allegedly true textual representation of a spoken utterance against a speech-

recognized textual representation of the same utterance. The alignment or pairing can occur on a word-by-word basis. Updates to a word database, which is a speech vocabulary or grammar used when speech-to-text converting utterances, can be automatically made based upon comparisons between the allegedly true textual representation and the speech-recognized textual representation.

Notably, during the alignment process, a link can be maintained between the speech-recognized textual representation and the speech utterance from which the textual representation was derived (note the claimed limitation of "corresponding speech segment" of claims 1 and 17, the "corresponding realization information" of claims 6 and 22, and the "combination of spoken audio and text" from claim 11). The linkage and audio information is needed to perform the update of the word database.

As already mentioned, Glickman is directed towards the problem of forced alignment that is different from the problem of speech recognition in that unlike speech recognition (were words said are unknown) in forced alignment, the text is known, but the time alignment of the text with the spoken words of the audio stream is unknown. (See column 1, lines 31-37 of Glickman for this teaching.) For the purposes of Glickman, any spoken segments of the acoustic model 325 used by the recognizer 302 that are not associated with text included within the text segment 310, would constitute undesirable "noise" that would be detrimental for aligning the text segment 310 with the audio segment 330 (Applicants make this observation based upon the teachings of Glickman taken as a whole, as supported by such passages as column 4, lines 57-65).

For this reason, Glickman teaches a method that "iterates over smaller and smaller unaligned segments with a more constrained vocabulary and language model" (column 4, lines 57-58) at each iteration. That is, a key principle of operation taught by Glickman that is instrumental for achieving the purpose for which Glickman is intended is that a recognition grammar used by the recognizer is to continuously and iteratively be reduced. In each iteration, a new segment of "anchored" but theretofore unaligned text is refined,

until either a desired termination condition is reached or the text and audio files are fully aligned (column 4, lines 36-38).

In other words, Glickman relies upon limiting the acoustic model 325 to those words contained within the voice and language model 320, where the entries in the voice and language model 320 originally contain all the words present within the text segment 310, and eventually contains only those words present between two anchors of an unaligned text portion located between the anchors. An alignment refining iteration can cause alignment weights (confidence scores) to be adjusted, as shown by the training data of FIG. 3, and as detailed between column 3, line 63 and column 4, line 13. Thus, "at each iteration, the vocabulary and language model 320 is rebuilt only from the words of that text segment, and only that particular V-LM 320 is used to recognize words in the current audio segment 330" (column 4, lines 44-47). Because of the above, not only does Glickman fail to explicitly or inherently teach the Applicants claimed invention, but actually teaches away from the claimed invention.

Further, Applicants claim that an update to a word database is based upon not only results (either non-aligned words from claim 1, matching word pairs from claim 6, or at least one word of the single word pair from claim 11) from an aligning step but also upon a corresponding segment of audio. Glickman provides no similar teachings regarding the utilization of an audio segment and a corresponding text segment. Instead, Glickman teaches that within a temporally organized audio stream, a series of markers can be placed. These markers are used to restrict a text list and to distinguish aligned audio segments from unaligned audio segments.

Applicants note that any modifications of Glickman designed to expand the acoustic model 325 based upon alignment information would not only contradict the explicit teachings of Glickman, but would also render Glickman unsatisfactory for its intended purpose (not permissible by MPEP 2143.01) and would change the principle of operation of Glickman (not permissible by MPEP 2143.01). Thus, Applicants do not

believe Glickman is a relevant prior art reference for purposes of the claimed application (as currently amended).

In summary, Applicants have amended claims and have shown how the claim amendments are supported by the specification. Applicants have also shown that Glickman fails to explicitly or inherently teach each claimed limitation (especially that of extending a word database) and have thereafter requested that the rejections to claims 1-28 be withdrawn. Applicants have then detailed how Glickman's teachings contradict those teachings present in the Applicants claimed invention in fundamental ways.

The Applicants believe that the application in its present form, including claims 1-28, is now in full condition for allowance, which action is respectfully requested. The Applicants request that the Examiner call the undersigned (direct line 954 759-8937) if clarification is needed on any matter within this Amendment, or if the Examiner believes a telephone interview would expedite the prosecution of the subject application to completion.

Respectfully submitted,

Date: 14 March 2005

Gregory A. Nelson, Registration No. 30,577
Brian K. Buchheit, Registration No. 52,667
Richard A. Hinson, Registration No. 47,652
AKERMAN SENTERFITT
Customer No. 40987
Post Office Box 3188
West Palm Beach, FL 33402-3188
Telephone: (561) 653-5000

{WP215417;3}